



CohortNet: Empowering Cohort Discovery for Interpretable Healthcare Analytics

Qingpeng Cai
National University of Singapore
Singapore
qingpeng@comp.nus.edu.sg

Kaiping Zheng
National University of Singapore
Singapore
kaiping@comp.nus.edu.sg

H. V. Jagadish
University of Michigan
Michigan, United States
jag@umich.edu

Beng Chin Ooi
National University of Singapore
Singapore
ooibc@comp.nus.edu.sg

James Yip
National University Health System
Singapore
james_yip@nuhs.edu.sg

ABSTRACT

Cohort studies are of significant importance in the field of healthcare analytics. However, existing methods typically involve manual, labor-intensive, and expert-driven pattern definitions or rely on simplistic clustering techniques that lack medical relevance. Automating cohort studies with interpretable patterns has great potential to facilitate healthcare analytics and data management but remains an unmet need in prior research efforts. In this paper, we present a cohort auto-discovery framework for interpretable healthcare analytics. It focuses on the effective identification, representation, and exploitation of cohorts characterized by medically meaningful patterns. In the framework, we propose CohortNet, a core model that can learn fine-grained patient representations by separately processing each feature, considering both individual feature trends and feature interactions at each time step. Subsequently, it employs K-Means in an adaptive manner to classify each feature into distinct states and a heuristic cohort exploration strategy to effectively discover substantial cohorts with concrete patterns. For each identified cohort, it learns comprehensive cohort representations with credible evidence through associated patient retrieval. Ultimately, given a new patient, CohortNet can leverage relevant cohorts with distinguished importance which can provide a more holistic understanding of the patient's conditions. Extensive experiments on three real-world datasets demonstrate that it consistently outperforms state-of-the-art approaches, resulting in improvements in AUC-PR scores ranging from 2.8% to 4.1%, and offers interpretable insights from diverse perspectives in a top-down fashion.

PVLDB Reference Format:

Qingpeng Cai, Kaiping Zheng, H. V. Jagadish, Beng Chin Ooi, and James Yip. CohortNet: Empowering Cohort Discovery for Interpretable Healthcare Analytics. PVLDB, 17(10): 2487 - 2500, 2024.
doi:10.14778/3675034.3675041

PVLDB Artifact Availability:

The source codes are available at <https://github.com/KimballCai/CohortNet>.

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit <https://creativecommons.org/licenses/by-nc-nd/4.0/> to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.
Proceedings of the VLDB Endowment, Vol. 17, No. 10 ISSN 2150-8097.
doi:10.14778/3675034.3675041

1 INTRODUCTION

Effective data management in healthcare [35, 42, 48–50] is pivotal for unlocking the vast potential inherent in electronic health records (EHRs). To facilitate clinicians in analyzing EHR data, it is crucial to explore cohorts that characterize patients sharing common conditions or responses to treatments. Studying the commonalities of cohorts among their associated patients enables clinicians to uncover valuable relationships and future outcomes. When assessing new patients, taking their relevant cohorts as auxiliary information empowers clinicians to gain a deeper understanding of their health condition, enhancing the quality of patient care. The significance of cohorts with insights has been confirmed in a multitude of studies, underscoring that the incorporation of cohort studies into healthcare analytics [16, 39, 45, 51, 52] and data management [17, 27, 29, 44] has the potential to facilitate healthcare research and patient care.

Cohort studies entail the identification of a specific group of patients, referred to as a “cohort,” based on a discernible “pattern”. Subsequently, this cohort is compared to other groups without such a pattern. Traditionally, these cohort-defining patterns are manually crafted by clinical experts, relying on features with specific value ranges (e.g., blood pressure > 140/90 mmHg) and are associated with outcomes of interest, such as disease progression and endpoints. Although these manual processes can yield medically meaningful insights, they demand extensive medical knowledge, and the subsequent identification of outcomes is both time- and resource-intensive. As such, incorporating automated approaches [28] for the auto-discovery of cohorts characterized by meaningful patterns is imperative for healthcare analytics.

Unfortunately, cohort discovery within the context of EHR data presents several challenges that have not been effectively addressed in prior studies. Cohort identification calls for the creation of interpretable and generalizable medical patterns, which should be defined at a granular level, e.g., feature level. Existing studies employ various approaches to form cohorts, including matching based on static features [23, 29, 34], rule-based formulations [47], and analyses of event trajectories [27]. However, these approaches are not effective for uncovering the underlying patterns in time-series features, as they fail to capture dynamic changes or feature interactions over time. Further, many features are recorded as numerical values, rendering the discovery of cohorts more challenging. To handle

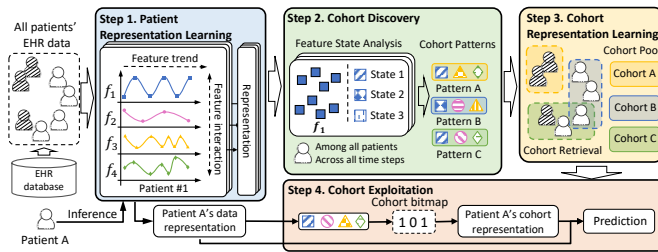


Figure 1: The systematic pipeline of cohort learning

such values, conventional approaches, such as discretization, not only require specific value range definitions but also constrain the search space for exploration, leading to suboptimal performance.

Some recent machine learning based approaches propose to employ clustering [30, 45, 46] or patient similarity [37, 53] to group patients into distinct “cohorts”. However, these approaches predominantly rely on latent representations at a coarse level, typically the patient level, lacking concrete and interpretable patterns. The absence of interpretability restricts their practical utility for clinicians. In some other studies [26, 27, 46], cohorts are created based on a small number of patients (e.g., a batch) for efficiency, which are not credible enough to convince clinicians.

To address above challenges, we propose an effective cohort discovery framework to identify cohorts with medically interpretable patterns, learn cohort representations, and utilize these cohorts in facilitating downstream analytics. Its pipeline sequentially executes four steps, as illustrated in Figure 1. At its core, we propose CohortNet, a novel model consisting of four modules for learning patients’ fine-grained EHR data representations for prediction and personalized cohort representations for calibration.

In Step 1, we introduce the *Multi-channel Feature Learning Module* in CohortNet to learn patients’ fine-grained representations. The module models each medical feature separately in a distinct channel to preserve features’ individuality while simultaneously capturing feature trends and interactions at each time step.

In Step 2, backed by the learned feature-level representations, *Cohort Discovery Module* classifies each feature into multiple distinct feature states (e.g., States s_1, s_2, s_3 for feature f_1). Thereafter, we propose a heuristic cohort exploration strategy that jointly considers feature states and interactions to uncover latent patterns among all patients and across all time steps. As a consequence, we are able to identify numerous concrete patterns over multivariate time series EHR data. For instance, Pattern A encompasses f_1 at s_1 , f_2 at s_1 , and f_3 at s_2 . In contrast to prior studies, feature states delineated in our patterns capture a broader spectrum of information, including value ranges, feature interactions, and others.

In Step 3, for each identified pattern, *Cohort Representation Learning Module* retrieves associated patients exhibiting this specific pattern, thereby constructing its respective cohort, exemplified by Cohort A derived from Pattern A. Through an exploration of the intrinsic information embedded within this cohort, the module derives the cohort’s comprehensive representations that demonstrate the associated patients’ commonality and thus, facilitate subsequent analytics tasks. For instance, as substantiated by diagnosis labels

and clinical experts, most of the patients within Cohort A ultimately manifest respiratory acidosis.

Lastly, **in Step 4**, by harnessing the learned credible cohorts in Step 3, *Cohort Exploitation Module* enables personalized analyses, concurrently enhancing performance and revealing medically significant insights. It initially identifies Patient A’s pertinent cohorts, referred to as the cohort bitmap. Subsequently, this module assesses the significance of these cohorts in learning the patient’s cohort representations and making predictions.

During training, our framework sequentially executes all four steps for identification, learning, and exploitation of cohorts. When inference on new patients, the framework performs only Steps 1 and 4 to derive the final decisions based on fine-grained data representations and personalized cohort representations. Our main contributions are summarized as follows:

- We propose a novel and general framework for effective cohort discovery without the infusion of external knowledge. Specifically, it can identify, represent, and utilize medically significant cohorts with concrete feature patterns, enhancing prediction and interpretation simultaneously. To the best of our knowledge, the auto-discovery of cohorts with interpretable patterns, despite its importance, is not attainable in the past.
- It features a systematic pipeline with four steps for healthcare analytics, centered around CohortNet’s four innovative modules tailored for each step. CohortNet is designed to learn feature-level representations through fine-grained modeling of feature trends and interactions, thereby facilitating subsequent feature state analysis. Through an exploration of the interrelationships among features, we devise a heuristic cohort exploration strategy for the auto-discovery of cohorts with concrete feature patterns. For each pattern, CohortNet retrieves its patients, forming its cohorts with comprehensive representations. Ultimately, by modeling patients’ individual data and identifying their associated cohorts with distinguished importance, CohortNet can provide personalized assessment with interpretable insights.
- We evaluate its effectiveness on three real-world EHR datasets. Extensive experimental results validate that, by taking into account discovered valuable cohorts, it consistently outperforms state-of-the-art baselines, with significant improvements in AUC-PR ranging from 2.8% to 4.1%. Meanwhile, when making decisions, it can offer medically interpretable insights from different perspectives that are consistent with findings in previous studies.

In the remainder of this paper, we review related work in Section 2, present our framework in Section 3, and delve into CohortNet with detailed descriptions of its modules. Following this, Section 5 provides a comprehensive discussion of its experimental results and the evaluation of interpretability. Finally, we conclude in Section 6.

2 RELATED WORK

EHR Data Modeling. The burgeoning of deep learning and its extensive utilization in healthcare analytics [14] and management [40] has spurred researchers to develop multifarious advanced deep learning models with libraries [28, 32] for modeling EHR data. Specifically, recurrent neural networks (RNN) such as long short-term memory (LSTM) [15] and gated recurrent units (GRU) [7] are widely adopted for handling time-series EHR data. Several studies

have designed specific mechanisms to improve EHR data modeling performance and interpretability. For instance, for the consideration of interpretation, RETAIN [5] employs two GRUs in reverse time order to distinguish attention at the visit and variable levels, while Dipole [22] adopts attention mechanisms to differentiate relationships among patients' visits. Besides, ConCare [24] embeds each medical feature separately and designs a multi-head self-attention mechanism to learn the interrelationships among medical features. From another perspective, some approaches focus on capturing specific concepts in EHR for boosted performance. T-LSTM [2] devises a time-aware LSTM to capture the irregular time intervals in EHRs, and StageNet [12] integrates stage modules to extract and utilize the information about disease stages. In contrast to the aforementioned studies that model patients' representations solely from individual EHR data, our framework can also discover medically meaningful cohorts from EHR data, learning overall information, and leverage these learned cohorts to enhance performance and interpretability.

Clustering. Clustering methods aim to group similar samples within a dataset based on similarity metrics. Approaches such as K-Means clustering [25], hierarchical clustering [20], and co-clustering [9] have demonstrated satisfactory performance across diverse applications. Unfortunately, their learned cluster results lack interpretable definitions, such as feature-based patterns, which hampers their acceptance by clinicians. Besides, most clustering methods fall short when analyzing EHR data due to their limited consideration of temporal sequences. While some studies have attempted to incorporate temporal relationships [3, 38, 41], they struggle to adequately address the complexities inherent in multivariate time series data from diverse patients - a challenge we aim to tackle in processing EHR data. Although these methods may perform well within specific datasets, their applicability to assessing new samples is non-trivial.

A few studies have attempted to facilitate deep learning tasks with these learned clusters. For instance, GRASP [46] uses the K-Means algorithm to group batches of patients' overall representations into several clusters and employs the K-Nearest neighbor (KNN) to utilize these clusters. PPN [45] clusters all patients' representations and selects the patients closest to centroids as typical patients, potentially deviating from centroids. However, both learned clusters and typical patients are identified and utilized via similarity scores (e.g., Euclidean distance) measured based on patients' learned representations, which lack explicit feature patterns and fine-grained interpretability. In contrast, we focus on discovering universal cohorts characterized by interpretable feature patterns.

Cohort Study. Cohort studies, as a cornerstone of observational research in medicine, typically require experts to manually define concrete patterns in statistics or relevant medical features [29, 31] to determine cohorts and their constituent patient populations. For example, authors in [43] use glucose to define groups and investigate their associated endpoints. However, crafting such explicit feature patterns could be time-consuming and labor-intensive.

To automate cohort studies, in [27], a heuristic sequence matching algorithm is proposed to derive cohort representation by aggregating patients' trajectories. However, this approach can be intricate and computationally intensive when generating representative cohorts. Further, it is primarily suitable for analyzing patients'

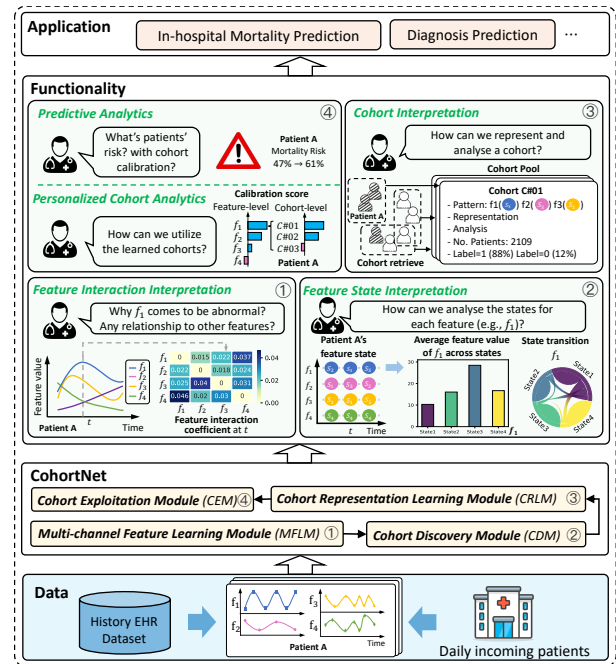


Figure 2: Our framework for the auto-discovery of cohorts in healthcare analytics with Patient A as a running example.

actions or events (such as disease occurrence or medication usage) and is not applicable to more complex numerical medical features. Besides, some studies attempt to identify similar patient groups by exclusively relying on static characteristics, such as gender and occupation. For example, RIM [34] and RB-GBDT [23] match identical non-time series features in structured data to identify similar patients. RuDi [47] categorizes patient groups by rules. It designs operators for features and converts knowledge from black-box models into rule-based student models. However, precise matching has restricted potential in learning cohorts for EHR data which comprises time-series medical features with numerical values.

Differing from these existing approaches, we propose CohortNet that employs deep learning techniques to automatically identify, learn, and exploit cohorts, facilitating clinical decision-making processes with boosted performance and interpretable medical insights.

3 FRAMEWORK

We shall use Patient A in Figure 2 as a running example to illustrate how the framework employs CohortNet on patients' EHR data to discover cohorts and make predictions, providing valuable functionalities for medical interpretation.

Data. Our framework initially collects comprehensive EHR data from hospital databases, including lab tests and vital signs, focusing on patients like Patient A. This dataset is pivotal for learning patients' data representations and discovering significant cohorts. **CohortNet.** At the heart of our framework lies the model CohortNet, which comprises four advanced modules: (a) *Multi-channel Feature Learning Module* (MFLM) utilizes patients' individual EHR data to learn informative representations by separately modeling

each feature while simultaneously capturing feature interactions and trends; (b) *Cohort Discovery Module* (CDM) first analyzes the states of features and then employs a heuristic cohort exploration strategy to unveil substantial cohorts characterized by specific feature patterns; (c) *Cohort Representation Learning Module* (CRLM) retrieves each cohort’s associated patients by utilizing its pattern and further analyzes its representations while extracting other essential information (e.g., label distributions); (d) *Cohort Exploitation Module* (CEM) identifies a patient’s relevant cohorts with differentiated importance and generates personalized cohort representations for calibration. More details about CohortNet are in Section 4.

Functionality. By prioritizing interpretability, we demonstrate our framework’s capability to yield medically interpretable insights.

- *Feature-level Interaction Interpretation.* When assessing Patient A’s abnormal medical features, such as f_1 , it is crucial to take into account its related features (e.g., f_2, f_3, f_4), allowing for a more accurate judgment of Patient A’s conditions (e.g., any complications). Supported by devised MFLM, our framework can explicitly model these interactions across time with differentiated importance, facilitating the analysis for each feature.
- *Feature State Interpretation.* In our framework, each medical feature will be grouped into distinct states through CDM to facilitate subsequent cohort discovery and exploitation processes. In Figure 2, each of Patient A’s features, such as f_1 , is classified into several distinct states, with each state inherently encapsulating unique meanings. For instance, states s_1, s_2 , and s_3 of f_1 are delineated by varying value ranges, transition pathways, and interactions, gaining a deeper understanding of the feature’s conditions. Hence, these feature states establish a robust interpretability foundation for cohort creation and subsequent steps.
- *Cohort Interpretation.* Through CDM, our model acquires the ability to discover an extensive cohort pool, where each cohort is characterized by a concrete pattern. In particular, each pattern consists of several medical features and their corresponding states. As an example depicted in Figure 2, the pattern for Cohort C#01 involves f_1 in s_3, f_2 in s_3, f_3 in s_1 . For each cohort, it deepens its understanding through further exploration of its associated patients. Hence, all cohorts learned in our framework are convincing with credible evidence (e.g., the number of associated patients, the label distribution).
- *Personalized Cohort Analytics.* As CRLM effectively learns significant cohorts, it is crucial to utilize these cohorts as references when providing personalized analytics. For instance, in the analysis of feature f_1 for Patient A, CEM identifies relevant cohorts (e.g., C#01, C#02, C#03, etc) tailored to Patient A’s changing conditions. Analyzing these cohorts to examine the calibration scores at both cohort and feature levels, CEM can offer a granular understanding of Patient A, facilitating more accurate predictions.
- *Predictive Analytics.* Our framework analyzes patient health conditions using quantified prediction scores to provide timely alerts for potential deterioration. By leveraging Patient A’s individual EHR data as the foundation (with a mortality risk of 47%) and calibrating with relevant cohorts (increasing risk to 61%), our framework can effectively and accurately predict outcomes, offering medically interpretable insights.

Table 1: Notations in our framework.

Notation	Definition
\mathbf{X}	EHR data
\mathbf{m}	masking vector for medical features
\mathbf{e}	feature embedding
\mathbf{u}	representation of feature interaction
\mathbf{v}	representation of feature trend
α	attention scores of a feature’s interactions
\mathbf{o}	fused representation at the feature level
\mathbf{h}	overall EHR data representation at the feature level
\mathbf{h}'	overall cohort representation at the feature level
$\tilde{\mathbf{h}}$	overall EHR data representation at the patient level
$\hat{\mathbf{h}}$	overall cohort representation at the patient level
s	feature state
ψ	pattern mask
η	cohort pattern
$C(\eta)$	cohort’s overall representation
ξ	cohort
$Pool(\xi)$	cohort pool
\mathbf{l}	labels
β	attention scores of a feature’s relevant cohorts
\mathbf{b}	personalized cohort bitmap
z	overall calibration score

Application. Our framework can tackle diverse downstream applications in healthcare analytics. In this paper, we focus on two extensively studied tasks, namely in-hospital mortality prediction and diagnosis prediction, to demonstrate how it facilitates clinicians in delivering accurate predictions with interpretability.

4 CohortNet

In this section, we delve into the detailed design of CohortNet.

EHR data. A patient’s EHR data comprises multivariate time series collected over time. Each time series denotes a medical feature that may have varying frequencies and lengths. We process each time series at regular intervals, resulting in each patient’s data being represented as sequences of medical features, denoted as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|F|}\}$. $\mathbf{x}_i \in R^T$ represents the sequence of the i -th medical feature over T time steps, and $|F|$ denotes the number of medical features. Furthermore, we introduce a masking vector $\mathbf{m} \in \{0, 1\}^{|F|}$ where $m_i = 0$ means that the i -th feature never presents in a patient’s EHR data. To lay the groundwork for our investigation, we first define several core concepts as below.

Definition 4.1 (Cohort). A cohort is defined as a tuple $\xi = \langle \eta, C(\eta) \rangle$ where η denotes the cohort pattern utilized for patient retrieval, and $C(\eta)$ signifies the cohort representation used in prediction tasks. A patient is classified into a cohort if and only if this patient manifests the specific cohort pattern at any time step.

Definition 4.2 (Cohort Pattern). Each cohort is characterized by a distinctive cohort pattern, comprising several medical features along with their respective states, formally defined as $\eta = \{(i, s_i)\}^d$,

where i is the feature index, s_i represents the state of the i -th feature, and d indicates the number of involved features in the pattern.

Definition 4.3 (Cohort Representation). Each cohort's representation, denoted as $C(\eta)$, is derived from the retrieved patients' representations within the cohort and is subsequently utilized in predictive analyses of patient conditions.

To facilitate automated cohort identification, learning, and exploitation, we outline an analytic task comprising four sequential steps: patient representation learning, cohort discovery, cohort representation learning, and cohort exploitation, with details below.

Definition 4.4 (Patient Representation Learning). Given a patient's EHR data \mathbf{X} , this step to learn a mapping function \mathcal{F} to generate the patient's overall representation $\tilde{\mathbf{h}}$ by summarizing feature-level representations \mathbf{h} . It also strives to encapsulate essential patient information, including attention scores α of feature interactions, derived from the patient's EHR data \mathbf{X} , i.e., $\tilde{\mathbf{h}}, \alpha = \mathcal{F}(\mathbf{X}, \mathbf{m})$.

Definition 4.5 (Cohort Discovery). For each feature, the step of cohort discovery initially utilizes a clustering algorithm to partition each feature into distinct states s . Subsequently, a cohort exploration strategy is employed in function \mathcal{G} to explore a cohort pool $Pool(\xi)$ considering interactions from learned attention scores α , where each cohort exhibits a specific pattern η as per Definition 4.2. This step is formalized as $Pool(\xi) = \mathcal{G}(\alpha, \mathbf{h})$.

Definition 4.6 (Cohort Representation Learning). Upon identifying a cohort pattern η , a representation learning algorithm \mathcal{C} is applied to retrieve all associated patients and learn the corresponding cohort representations (as per Definition 4.3) from these patients.

Definition 4.7 (Cohort Exploitation). This step of cohort exploitation examines a patient's EHR data \mathbf{X} alongside the cohort pool $Pool(\xi)$. It begins by identifying each patient's pertinent cohorts through a cohort bitmap \mathbf{b} , subsequently deriving the patient's personalized cohort representation $\hat{\mathbf{h}}$. The final prediction \tilde{y} , computed using the function \mathcal{U} , integrates the patient's individual data representation $\tilde{\mathbf{h}}$ with the patient's cohort representation $\hat{\mathbf{h}}$ as calibration. This function is denoted as $\tilde{y} = \mathcal{U}(\tilde{\mathbf{h}}, \hat{\mathbf{h}})$.

Our overall objective is to learn a cohort pool $Pool(\xi)$ while concurrently learning the mapping functions \mathcal{F} , \mathcal{G} , \mathcal{C} , and \mathcal{U} to accurately predict the outcome \tilde{y} . To streamline the presentation, we frame our prediction task as a binary classification without loss of generality, where $\tilde{y} \in \{0, 1\}$. We summarize the notations used throughout the paper in Table 1. To enhance readability, we omit the superscript t when analyzing a single time step.

4.1 Multi-channel Feature Learning Module

Some prior methods for processing EHR data, such as Channel-wise LSTM [13] and ConCare [24], conceptualize a patient as a composite of her/his medical features. They employ separate models to encode the raw values of each medical feature, thereby preserving the individuality of each feature. However, these raw values are insufficient for modeling complicated feature conditions or feature interactions [4]. In contrast to these methods, MFLM, as depicted in Figure 3, attains fine-grained representations of EHR data at the feature level, learning feature trends and interactions across time.

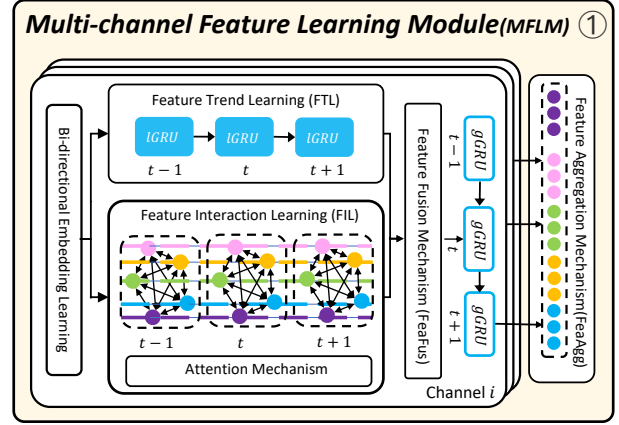


Figure 3: The Multi-channel Feature Learning Module.

To preserve each feature's individuality, MFLM employs multiple channels to learn a patient's overall representation, with each channel focusing on a particular feature. In each channel, we first convert each medical feature's numerical values into meaningful embedding vectors \mathbf{e} via a Bi-directional Embedding Learning (BiEL) mechanism [4] that has been validated to be beneficial for embedding numerical features with a predefined bounds a and b :

$$\mathbf{e}_i = \begin{cases} \frac{1}{b-a} (\mathbf{V}_i^a (x'_i - a) + \mathbf{V}_i^b (b - x'_i)), & \text{if } \mathbf{m}_i = 1, \\ \mathbf{V}_i^m, & \text{if } \mathbf{m}_i = 0. \end{cases} \quad (1)$$

where $\mathbf{V}^a, \mathbf{V}^b, \mathbf{V}^m \in \mathbb{R}^{|F| \times d_e}$ are embedding matrices that convert x'_i into a lower-dimensional feature representation, and $|F|$ and d_e denote the number of features and dimension of \mathbf{e} . This mechanism preserves the advantages of a linear embedding mechanism while simultaneously controlling the embedding scale and enhancing the embedding quality. Besides, we also integrate the Feature Interaction Learning (FIL) mechanism [4] which enables the capture of explicit feature interactions (\mathbf{u}) and their corresponding attention scores (α) using fewer training parameters. Taking the i -th feature for illustration, the FIL mechanism models its interactions with all the other features at each time step:

$$(\mathbf{u}_{i,1}, \alpha_{i,1}), (\mathbf{u}_{i,2}, \alpha_{i,2}), \dots, (\mathbf{u}_{i,|F|}, \alpha_{i,|F|}) = \text{FIL}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{|F|}) \quad (2)$$

By learning these fine-grained feature interactions, CohortNet can capture diverse valuable feature patterns, facilitating the subsequent cohort auto-discovery processes. Further, the temporal trends of medical features also are essential in the analysis. To capture such trends, we propose a Feature Trend Learning (FTL) mechanism that utilizes separate local GRUs (IGRU) to learn the temporal dynamics of each feature. Specifically, given i -th feature's embeddings, FTL leverages an individual $IGRU_i$ to model its temporal behaviors \mathbf{v} :

$$\mathbf{v}_i^1, \mathbf{v}_i^2, \dots, \mathbf{v}_i^t = IGRU_i(\mathbf{e}_i^1, \mathbf{e}_i^2, \dots, \mathbf{e}_i^t), t = \{1, 2, \dots, T\} \quad (3)$$

Next, to derive the overall representation of a certain feature \mathbf{o} , we further devise a Feature Fusion (FeaFus) mechanism in MFLM to incorporate feature embedding \mathbf{e} , feature interactions \mathbf{u} , and feature trend \mathbf{v} through a multilayer perceptron.

$$\mathbf{o}_i^t = \text{FeaFus}(\mathbf{e}_i^t, \mathbf{u}_i^t, \mathbf{v}_i^t) \quad (4)$$

Through this adaptive fusion mechanism, we derive a fused representation \mathbf{o} for each medical feature with reduced dimensionality, facilitating computations for the following cohort discovery.

Subsequently, to learn each feature’s overall EHR representation \mathbf{h} , MFLM employs separate global GRUs (gGRUs) in each channel to maintain the learning focus:

$$\mathbf{h}_i^1, \mathbf{h}_i^2, \dots, \mathbf{h}_i^t = gGRU_i(\mathbf{o}_i^1, \mathbf{o}_i^2, \dots, \mathbf{o}_i^t), t = \{1, 2, \dots, T\} \quad (5)$$

Finally, we propose a feature aggregation (*FeaAgg*) mechanism that integrates the representations of all medical features learned in their respective channels into a holistic representation $\tilde{\mathbf{h}}^t$.

$$\tilde{\mathbf{h}}^t = FeaAgg(\mathbf{h}_1^t, \mathbf{h}_2^t, \dots, \mathbf{h}_{|F|}^t) \quad (6)$$

In *FeaAgg*, we compress and concatenate all feature-level representations to form the overall patient-level representation, allowing for a fine-grained understanding of the patient’s feature conditions.

In CohortNet, MFLM serves as a solid foundation, facilitating following cohort discovery. Besides, it supports fine-grained feature-level interaction interpretation, depicting diverse abnormal feature conditions. In this step, we compute the feature embedding in \mathbf{e} in $\mathcal{O}(|F| \times T \times d_e)$, the feature trend \mathbf{v} in $\mathcal{O}(|F| \times T \times (3d_e d_t + 3d_t^2))$, the feature interaction \mathbf{u} in $\mathcal{O}(|F|^2 \times d_e)$, the fused representation \mathbf{o} in $\mathcal{O}(|F| \times (2d_e + d_t)d_o)$, the overall feature representation \mathbf{h} in $\mathcal{O}(|F| \times T(3d_o d_h + 3d_h^2))$, and the overall patient representation $\tilde{\mathbf{h}}$ in $\mathcal{O}(|F| \times (d_o \times d_p))$ where d_e , d_t , d_o , d_h , and d_p denote the dimensions of \mathbf{e} , \mathbf{v} , \mathbf{o} , \mathbf{h} , and $\tilde{\mathbf{h}}$, respectively. Overall, we achieve a complexity of $\mathcal{O}(|F| \times T \times (d_o d_h + d_h^2))$ in this step.

4.2 Cohort Discovery Module

In medical studies [29, 31, 43], clinicians traditionally engage in manual pattern identification to classify patients into distinct cohorts, which is time-consuming and labor-intensive, requiring extensive domain knowledge. Further, in [27], formalizing cohorts as trajectories of patient events, particularly in EHR data containing numerical values, poses significant challenges. While K-Means clustering employed in [45, 46] can group patients based on their learned overall representations, the centroids learned in clusters are limited in interpretability. To overcome these constraints, we introduce CDM to automate the identification of universal cohorts exhibiting concrete feature patterns across multiple sequences of EHR data. The overview of this module is shown in Figure 4.

On account of the necessity of leveraging specific features in patterns, CDM begins by classifying each feature into distinct states based on the learned representations \mathbf{o} . To achieve this, for each feature, CDM gathers all its possible representations from all samples at all time steps and then employs a clustering algorithm, denoted as *Cluster*, to analyze and identify the feature into distinct states.

$$\{s_0, s_1, s_2, \dots, s_k\} = Cluster(\mathbf{o}_i, m_i, k) \quad (7)$$

Compared with alternative clustering techniques such as hierarchical clustering and co-clustering, we ultimately select K-Means in this module due to its superior efficiency, and the centroids learned in K-Means are easier to apply when assessing new patients. In detail, we employ the K-Means algorithm in an adaptive manner to identify k distinct states s_i for the i -th medical feature. Different states of the same medical feature reflect distinct conditions, which can be approximately characterized by the average feature values

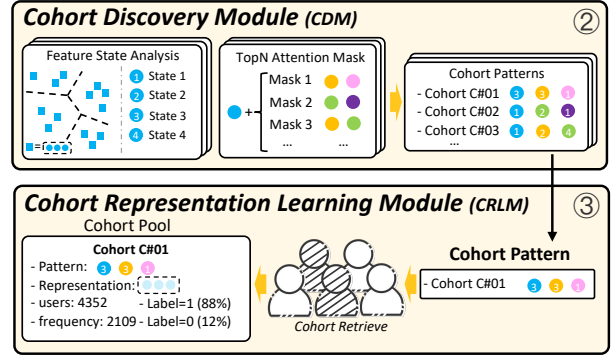


Figure 4: The Cohort Discovery Module and the Cohort Representation Learning Module.

of each state, feature interactions, and state transition pathways. As part of our algorithm, we particularly classify missing features (i.e., $m_i = 0$) into a distinct state s_0 since the absence of a feature directly represents a particular condition.

Relying solely on a feature’s state to represent the medical pattern can be inaccurate, while analyzing all medical features introduces the curse of dimensionality. More specifically, when extracting n' medical features from EHR data and clustering them into k clusters for each feature, the search space expands exponentially to $\mathcal{O}(k^{n'})$, leading to a increased computational complexity and a higher risk of overfitting in pattern discovery.

In practical medical scenarios, when analyzing the feature abnormality, clinicians jointly consider several other highly related features along with their values to achieve a more accurate assessment. Inspired by it, we address the cohort pattern exploration challenges by introducing a heuristic cohort exploration strategy in CDM that can leverage feature interactions as guidance without the infusion of external medical knowledge. Specifically, it identify substantial cohort patterns by analyzing both feature interactions (i.e., α learned in MFLM) and their corresponding feature states.

When processing i -th medical feature, CDM analyzes its feature interactions by transforming attention scores α into an attention-based pattern mask $\psi_i = \{0, 1\}^{|F|}$:

$$\psi_i = topN(\alpha_i, n) + onehot(i) \quad (8)$$

where $topN(\cdot, \cdot)$ generates a binary vector identifying n relevant medical features with the highest attention scores. Additionally, $onehot(\cdot)$ constructs another binary vector where only the i -th dimension is set to one. In CDM, each medical pattern η is derived from the combination of both the attention-based mask ψ and the feature states \mathbf{s} , $\eta_i = \mathbf{s} \odot \psi_i$ where \odot denotes the element-wise product. With such designed cohort exploration strategy employed in CDM, CohortNet can effectively find meaningful medical patterns without covering the whole search space, and each pattern η involves d_ψ features, $\|\psi_i\|_1 = d_\psi = n + 1$. In this step, we cluster feature states in $\mathcal{O}(|F| \times (|P| \times T)kd_h I)$ and analyze patterns in $\mathcal{O}(|F| \times (|P| \times T) \times d_\psi \log((|P| \times T) \times d_\psi))$, where $|P|$ denotes the number of patients used for discovering cohorts. k and I refer to the number of clusters and iterations used in K-Means. Thus, the complexity of CDM is $\mathcal{O}(|F| \times (|P| \times T) \times d_\psi \log((|P| \times T) \times d_\psi))$.

4.3 Cohort Representation Learning Module

In prior studies, cohorts are delineated through various means, encompassing patient event trajectories [27], hand-crafted medical definitions [29, 31, 43], and K-Means centroids [46]. However, these approaches either lack compatibility with deep learning techniques or encounter challenges regarding interpretability. In contrast, our framework not only discerns cohorts with medically interpretable feature patterns but also furnishes informative cohort representations for predictive tasks. Specifically, derived from its pattern η_i^q (the q -th pattern for the i -th feature), the Cohort Representation Learning Module (as shown in the bottom of Figure 4) retrieves the patients who exhibit this particular pattern and learns representations from these patients' commonalities.

In CohortNet, a patient belongs to the cohort ξ_i^q if and only if all states of the involved features in its pattern η_i^q exactly match the patient's feature states at a certain time step. With the retrieved patients, we derive the cohort's latent representation $C(\eta_i^q)$ by learning from these patients' representations and labels:

$$C(\eta_i^q) = \left[\frac{1}{|\xi_i^q|} \sum_p h_i^p; \mathbf{l}_i^q \right], p \in \xi_i^q \quad (9)$$

where p is a specific patient in the cohort ξ_i^q , and $[\cdot; \cdot]$ represents a concatenation function; \mathbf{l}_i^q refers to the distribution of patients' task-relevant labels (e.g., mortality labels and diagnosis labels) and task-irrelevant labels (e.g., frequencies and demographic distributions) within this cohort. Hence, a cohort in CohortNet comprises both the feature-based concrete medical pattern and the general representation, denoted as $\xi_i^q = \langle \eta_i^q, C(\eta_i^q) \rangle$. Such an in-depth understanding of cohorts enables our framework to deliver interpretable insights when leveraging these cohorts for personalized analysis. With numerous patterns discovered in CDM, our CRLM will create a substantial cohort pool, denoted as $Pool(\xi) = \{\xi_i^1, \xi_i^2, \dots, \xi_i^q, \dots\} \in \mathbb{R}^{|F|*|C|}$ where $|C| = |C_1| + |C_2| + \dots + |C_{|F|}|$, and $|C_i|$ denotes the number of cohorts discovered for i -th medical feature. Overall, this step's complexity is $\mathcal{O}(|F| \times (|P| \times T) \times (d_h + |L|) \times |C|)$ where $|L|$ and $|C|$ denote the size of used labels and learned cohorts.

Additionally, we integrate several filters to narrow down the search space and mitigate the computation complexity, e.g., the sample frequency filter, to identify and exclude medical patterns that occur infrequently in the training samples, since low frequencies result in insufficient evidence to support these cohorts' credibility.

4.4 Cohort Exploitation Module

The acquired cohorts should be readily applicable to new patients efficiently and effectively. However, the cohorts learned in earlier work [46] have been limited by their applicability to specific patient populations. Furthermore, cohorts in [27] require substantial computational resources due to reliance on sequence matching methods. In CohortNet, CEM can efficiently leverage these credible and universal cohorts as supplementary information when assessing individual patients' feature conditions. That is, it takes the cohort pool $Pool(\xi)$ as extensive knowledge and indexes the patient's relevant cohort with differentiated importance. As a result, CohortNet not only improves the performance of downstream prediction tasks but also provides clinicians with interpretable cohort-based insights.

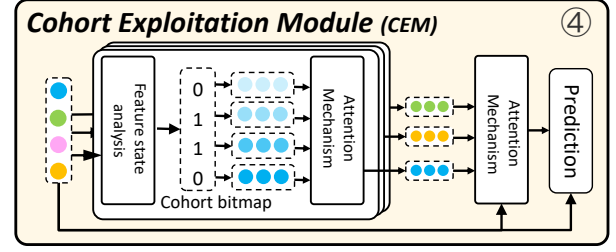


Figure 5: The Cohort Exploitation Module.

As shown in Figure 5, CDM first examines patients' feature states s at each time step and subsequently analyze cohort bitmap $\mathbf{b}_i = \{0, 1\}^{|C_i|}$ when analyzing i -th medical feature:

$$\mathbf{b}_i^q = 1, \text{ if } \exists s_i^t \odot \psi_i^q = \eta_i^q, t = 1, 2, \dots, T \quad (10)$$

where ψ_i^q is the pattern mask indicating the involved features in the pattern η_i^q . When integrating these cohorts to facilitate the evaluation of the i -th feature, it is crucial to distinguish the importance (denoted as β_i^q) of the involved cohorts. To address this, we devise an attention mechanism that discerns the varying significance of different cohorts in relation to the target prediction task:

$$\beta_i^q = (\mathbf{W}_Q \cdot \mathbf{h}_i^T) \cdot (\mathbf{W}_K \cdot C(\eta_i^q)), q = 1, 2, \dots, |C_i| \quad (11)$$

$$\beta_i^1, \beta_i^2, \dots, \beta_i^{|C_i|} = \text{softmax}(\beta_i^1, \beta_i^2, \dots, \beta_i^{|C_i|}) \quad (12)$$

$$\mathbf{h}_i' = \sum_q \beta_i^q (\mathbf{W}_V \cdot C(\eta_i^q)) \quad (13)$$

where \mathbf{W}_Q , \mathbf{W}_K , and \mathbf{W}_V are trainable weights, and $C(\eta_i^q)$ denotes the cohort representation derived in Equation 9. CEM then summarizes a feature's cohort representations by weighing them with the learned attention scores β , to derive the feature's cohort representation \mathbf{h}_i' . Similar to MFLM, we concatenate all features' cohort representations and generate the patient's overall cohort-related representation $\tilde{\mathbf{h}} = [\mathbf{h}_1'; \mathbf{h}_2'; \dots; \mathbf{h}_{|F|}']$.

Finally, based on a patient's overall data representation $\tilde{\mathbf{h}}$ learned in MFLM and the patient's individual overall cohort representation $\tilde{\mathbf{h}}$ learned in CEM, we conduct a binary prediction task via:

$$\tilde{y} = \sigma(\mathbf{w}^p \cdot \tilde{\mathbf{h}} + b^p + \mathbf{w}^c \cdot \tilde{\mathbf{h}}) \quad (14)$$

where \mathbf{w}^p , b^p , \mathbf{w}^c are trainable weights, and $\sigma(\cdot)$ is the sigmoid function. When deriving the final prediction via Equation 14, we can re-express $\mathbf{w}^c \cdot \tilde{\mathbf{h}}$ as a cohort-related overall calibration score z :

$$z = \mathbf{w}^c \cdot \tilde{\mathbf{h}} \quad (15)$$

$$= \sum_i \underbrace{\mathbf{w}_i^c \cdot \mathbf{h}_i'}_{\text{feature-level calibration score}} \quad (16)$$

$$= \sum_i \sum_q \underbrace{\mathbf{w}_i^c \cdot \beta_i^q (\mathbf{W}_V \cdot C(\eta_i^q))}_{\text{cohort-level calibration score}} \quad (17)$$

where the overall calibration score z can be decomposed into feature-level calibration scores, which can further be subdivided into cohort-level calibration scores. Consequently, the complexity of cohort exploitation in this module is $\mathcal{O}(|F| \times T \times |C| \times (d_h + |L|) + |F| \times (2d_h + |L|)) \sim \mathcal{O}(|F| \times T \times |C| \times (d_h + |L|))$.

Our framework is capable of supporting diverse tasks, and we exemplify it by the binary classification task, where the binary cross-entropy loss function is employed:

$$\mathcal{L} = -\frac{1}{B} \sum_{j=1}^B y_j \log(\tilde{y}_j) + (1 - y_j) \log(1 - \tilde{y}_j) \quad (18)$$

y_j and \tilde{y}_j represent the ground truth labels and prediction labels, respectively, and B denotes the batch size.

4.5 Overall Complexity

We train our framework with all the four aforementioned steps: learning patients’ fine-grained representations via MFLM in $\mathcal{O}(|F| \times T \times (d_o d_h + d_h^2))$, discovering diverse cohort patterns via CDM in $\mathcal{O}(|F| \times (|P| \times T) \times d_\psi \log((|P| \times T) \times d_\psi))$, representing cohorts via CRLM in $\mathcal{O}(|F| \times (|P| \times T) \times (d_h + |L|) \times |C|)$, and utilizing these cohorts via CEM in $\mathcal{O}(|F| \times T \times |C| \times (d_h + |L|))$. Therefore, the overall computational cost in the training phase is $\mathcal{O}(|F| \times (|P| \times T) \times (d_h + |L|) \times |C|)$. However, when making inferences on new patients, only the first and last steps are required, leading to an overall complexity of $\mathcal{O}(|F| \times T \times |C| \times (d_h + |L|))$.

5 EXPERIMENTS

5.1 Experimental Setup

We evaluate the effectiveness of our framework on three real-world benchmark EHR datasets: MIMIC-III, MIMIC-IV, and eICU. We perform the in-hospital mortality prediction, formulated as a binary classification task, on the first two datasets and conduct diagnosis prediction as a multi-label classification task on the third dataset.

MIMIC-III Dataset [19] is a publicly available ICU dataset collected at Beth Israel Deaconess Medical Center spanning from 2001 to 2012. We first sample 21,139 admissions from this dataset based on patients’ demographic data and EHR data as suggested by [13] and subsequently extract 63 clinically aggregated time-series medical features, related to vital signs and lab tests.

MIMIC-IV Dataset [18], a public contemporary EHR dataset released in 2023, provides valuable insights into a decade of admissions ranging from 2008 to 2019. Distinguished from MIMIC-III, this dataset comprises more recent data and a larger sample size. In particular, we extract 35,122 admissions from this dataset, with a focus on the top 70 most frequently occurring time-series features.

eICU Dataset [33] is collected from many critical care units across the contiguous United States, encompassing patients admitted to these units during 2014 and 2015. In our experiments, we sample 41,547 admissions with 25 diagnosis labels as suggested by [36], where each sample encompasses the 67 most frequently occurring features, which include 48 lab tests and 19 chart events.

In these datasets, each sample contains medical features recorded during the initial 48 hours following admission to the intensive care unit (ICU), and all features are applied a mean-std standardization.

Evaluation Metrics. We compare CohortNet with baseline models using the following metrics: the area under the receiver operator characteristic curve (AUC-ROC), the area under the precision-recall curve (AUC-PR), and the F1-score. Among these metrics, AUC-PR is the primary metric because it is the most informative score when

handling a highly imbalanced dataset [6, 8, 46]. We divide the samples into 80%:10%:10% for training, validation, and testing.

Implementation Details. Our experiments are conducted on a server with 48 Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz and a GeForce RTX 3090 GPU. For model training, we employ the Adam optimizer [21] and set the learning rate to $1e-3$.

Baselines. We compare CohortNet with the baselines below.

- **LSTM** [15] is an RNN-based model for time-series analytics.
- **GRU** [7] is another RNN-based time-series model but requires fewer parameters than LSTM.
- **RETAIN** [5] utilizes two levels of GRU in the reverse time order to differentiate the importance of visits and variables.
- **Dipole** [22] adopts a bidirectional GRU and devises attention mechanisms to calculate the relationships among time steps.
- **StageNet** [12] models disease progression stages and incorporates them into learning disease progression patterns.
- **T-LSTM** [2] designs a time decay mechanism to handle irregular time intervals in EHRs.
- **ConCare** [24] embeds each time-series medical feature separately and employs a self-attention model to learn the relationships among these features.
- **GRASP** [46] relies on a backbone model to learn patients’ general representations, uses K-Means to find a group of similar patients, and applies K-NN to integrate the groups’ information.
- **PPN** [45] identifies typical patients to serve as prototypes and leverages these prototypes by calculating similarity metrics when assessing new patients.

Moreover, we conduct an ablation study to assess the influence of our learned cohorts by comparing it with the following baselines:

- **CohortNet w/o c** removes all cohort-related modules, namely CDM, CRLM, and CEM. This evaluation aims to validate the impact of the learned cohorts on the overall performance.
- **CohortNet w c-** employs K-Means to cluster patients’ overall representations directly in CDM and then identifies patients’ relevant cohorts directly via K-Means in CEM. The comparison demonstrates the necessity of defining cohorts with concrete patterns at the feature level.

5.2 Main Results

The experimental results of CohortNet compared to baseline models on the three datasets are illustrated in Figure 6. Among the baselines, RETAIN exhibits inferior performance, as it prioritizes interpretability over performance. Dipole, StageNet, T-LSTM, and ConCare generally outperform the standard time-series models LSTM and GRU due to their elaborate mechanisms tailored for capturing specific characteristics in EHRs. GRASP and PPN enhance predictive performance by integrating representations of similar or typical patients. In comparison to the best-performing baseline models, CohortNet consistently achieves improved performance in all metrics, demonstrating substantial improvements in AUC-PR, with increases of 3.5%, 2.8%, and 4.1% on the MIMIC-III, MIMIC-IV, and eICU datasets, respectively. Furthermore, CohortNet also achieves higher F1 scores by over 5% on the in-hospital mortality prediction task. The superior performance of CohortNet stems from its capability of fine-grained feature-level representation learning and effectiveness of cohort discovery, learning, and exploitation.

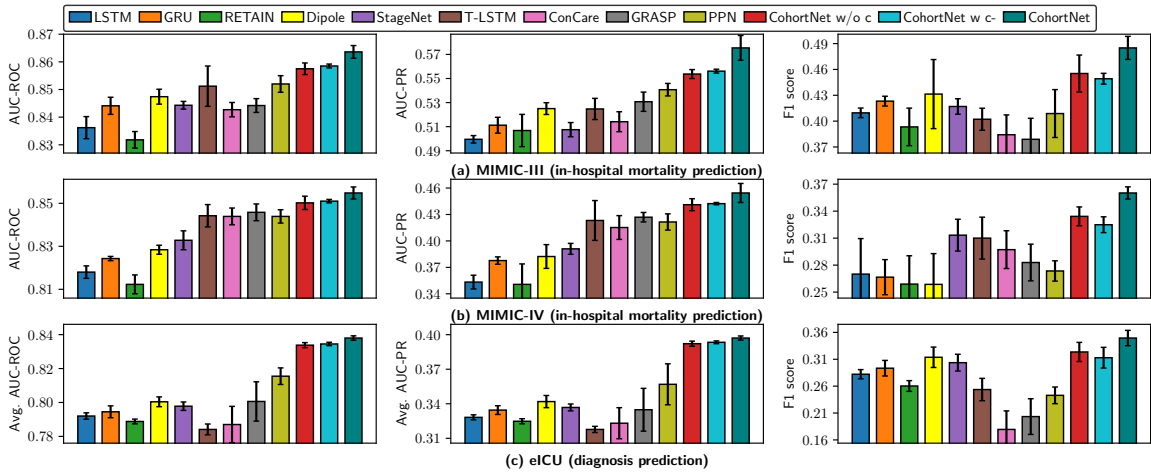


Figure 6: Experimental results for in-hospital mortality prediction on the MIMIC-III and MIMIC-IV datasets and diagnosis prediction on the eICU dataset in terms of the three metrics.

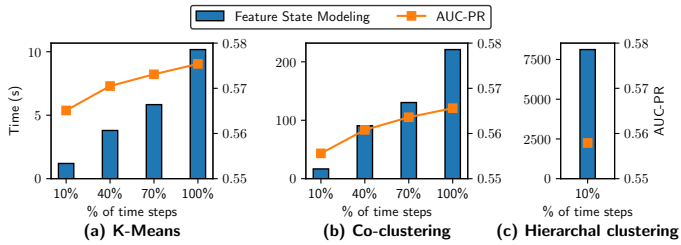


Figure 7: Efficiency and performance analysis of different clustering techniques for feature state modeling in CDM.

5.3 Ablation Study

We further conduct an ablation study to compare the performance of CohortNet and its several variants. As shown in Figure 6, CohortNet w/o c, which includes only MFLM, surpasses the other baselines, confirming the effectiveness of MFLM in fine-grained feature-level representation learning by integrating feature interactions and feature trends. The comparison against this variant underscores the efficacy of CohortNet, particularly in discovering significant cohorts defined by concrete feature patterns in CDM, further learned in CRLM, and ultimately exploited in CEM. This comprehensive pipeline in CohortNet leads to more accurate predictions, as indicated by improved F1 scores. Furthermore, CohortNet w c demonstrates merely marginal improvement, indicating that directly clustering patients’ overall representations cannot capture sufficient information for these prediction tasks. In contrast, CohortNet achieves substantial improvements, emphasizing the necessity of fine-grained cohort learning at the feature level.

To justify the adoption of K-Means for feature state modeling within the Cohort Discovery Module, we conduct a comparative analysis against two alternative clustering techniques: co-clustering [9] and hierarchical clustering [20], which both group data into predefined numbers of clusters. To mitigate the issue of time complexity in clustering, we sample patient representations

at varying sampling ratios of time step. As illustrated in Figure 7, co-clustering not only incurs greater time consumption than K-Means but also yields inferior performance in terms of AUC-PR. Hierarchical clustering, meanwhile, demonstrates prohibitive time consumption when modeling just 10% of time steps and suffers from memory exhaustion issues when handling larger volumes of data. Further, both co-clustering and hierarchical clustering require extra computations to determine cluster centroids for evaluating new patients, while K-Means can streamline this process using learned centroids. K-Means also achieves better efficiency and effectiveness which confirms the its usage in our Cohort Discovery Module.

5.4 Interpretation Study

In this section, we elucidate how our framework interprets predictions with learned cohorts, offering medically significant insights from different perspectives in a top-down fashion, as elaborated in Section 3. Figure 8(a) showcases Patient A’s standardized data for several critical features, with more details in the caption.

5.4.1 Predictive Analytics. We start our exploration of its predictive analytics by examining Patient A. Notably, we observe that our proposed Multi-channel Feature Learning Module predicts Patient A’s mortality probability to be 47%, based solely on Patient A’s EHR data representation \hat{h} . However, as shown in Figure 8(b), upon identifying and leveraging Patient A’s relevant cohorts, our entire framework increases this predicted probability to 61%.

5.4.2 Personalized Cohort Analytics. The changes in final predicted probabilities result from the examination of feature-level calibration scores, which are calculated in Eq.16, representing the influence of each feature on the final prediction. We illustrate these scores in Figure 8(c), with positive scores depicted in blue and negative scores in pink. Specifically, RR and BUN exhibit relatively higher feature-level calibration scores, thereby contributing to positive effects, while some other features (e.g., PIP) exert negative influences.

Table 2: Statistics of cohorts w.r.t RR.

Cohort	Frequency	Patients	Pos-Rate	Cohort Pattern
C#01	472	125	36.8%	RR(S3↑); BUN(S2↑); PCO2(S7↑)
C#02	12519	2188	29.3%	RR(S3↑); ALT(S3↓); AST(S2↓)
C#03	2909	1019	16.0%	RR(S3↑); HCO3(S3↑); PCO2(S7↑)
C#04	35753	12858	12.1%	RR(S2-); HR(S7-); HCO3(S2-)

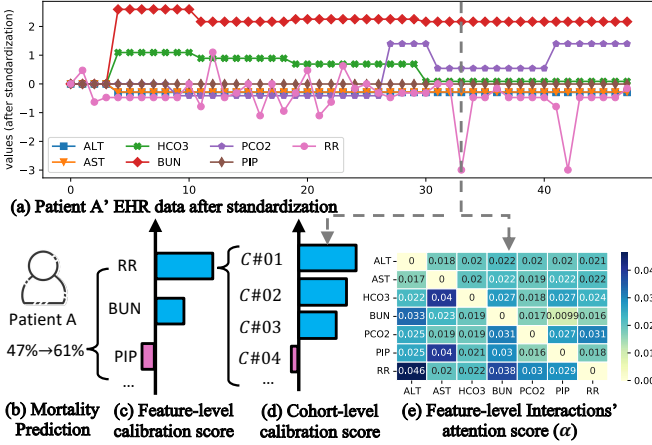


Figure 8: An example of how our framework analyzes Patient A with interpretable insights at different levels. Involved features include the respiratory rate (RR), alkaline phosphate (ALT), aspartate aminotransferase (AST), bicarbonate (HCO3), blood urea nitrogen (BUN), partial pressure of carbon dioxide (PCO2), peak inspiratory pressure (PIP).

Each feature’s feature-level calibration score is aggregated from the cohort-level calibration scores of its associated cohorts, as derived in Eq. 17. When analyzing the feature RR in Patient A, it identifies several relevant cohorts by evaluating features’ states at different time steps. For instance, Cohort C#01 is identified in the 34th hour via Eq. 10. As shown in Figure 8(d), delving into the feature RR, we discover some crucial cohorts with different cohort-level calibration scores. Certain cohorts (e.g., Cohorts C#01, C#02, and C#03) contribute to a higher risk of mortality, whereas others (e.g., Cohort C#04) might not exhibit this tendency. The details of all four relevant cohorts are presented in Table 2, including the number of associated patients (Patients) and the rate of patients with positive labels, i.e., patients’ mortality rate in this cohort (Pos-Rate).

5.4.3 Cohort Interpretation. Backed by the Cohort Discovery Module, each identified cohort has a distinct pattern comprised of multiple features along with their corresponding feature states. For instance, Cohort C#03 is characterized by the pattern where RR is in S3 (with lower values↓), HCO3 is in S3 (with higher values↑), and PCO2 is in S7 ↑. More details about the feature states will be introduced in Section 5.4.4. This cohort is crucial for analyzing the reason behind Patient A’s lower RR level and deteriorating health condition, as patients with such a pattern may be experiencing respiratory acidosis. Respiratory acidosis [10, 11] is a condition characterized by an elevation in PCO2 in the blood. Typically, it

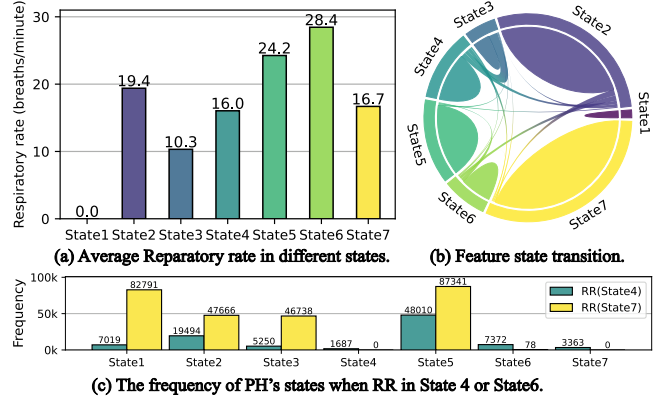


Figure 9: Feature state study in RR.

occurs when there is an accumulation of carbon dioxide in the body due to inadequate breathing or impaired gas exchange in the lungs. The low respiratory rate observed in the pattern suggests that patients are not breathing sufficiently to eliminate carbon dioxide effectively, resulting in elevated PCO2 levels. In response, the body compensates by increasing the concentration of HCO3 to maintain acid-base balance. In a nutshell, this cohort’s characteristics align with medical knowledge surrounding respiratory acidosis, which further validates its interpretability in learned cohorts.

Considering the complexity of a patient’s health condition, a patient may be associated with multiple cohorts that depict disease progression. As shown in Figure 8(d), Patient A is also associated with Cohort C#02 and C#01, which holds greater importance than C#03. Among them, Cohort C#01 holds Patient A’s highest cohort-level calibration score. Despite representing only 125 patients, this cohort exhibits a significantly higher mortality rate (36.8%) compared to other cohorts. Its cohort pattern features elevated BUN in S2 ↑, indicating increased blood nitrogen from urea, which is regulated by the liver and excreted by the kidneys. Respiratory acidosis, affecting kidney function, may consequently alter BUN levels. Existing medical studies [1, 10] explore the renal response to respiratory acidosis by analyzing relevant features like sodium and chloride. Given this context, it is promising to investigate this cohort, as abnormal features in its pattern may reveal possible interrelationships and inspire potential medical research advancements.

CohortNet also identifies several common cohorts such as Cohort C#04, which encompasses almost two-thirds of the patients in our training dataset. All features in its cohort pattern remain at a normal level (−), reflecting the patients’ normal conditions. Hence, C#04 demonstrates a slight negative influence on the final prediction when compared to the other cohorts with abnormal patterns. By jointly considering multiple relevant cohorts with varied importance, CohortNet can effectively facilitate clinicians with better assessments of the patient’s condition.

5.4.4 Feature State Interpretation. In Cohort Discovery Module, each feature is classified into k distinct states, and we take RR as an example to interpret feature states from three perspectives.

Firstly, the states of features generally correspond to varied medical values in EHRs. To account for this, we calculate the average

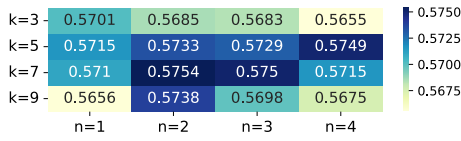


Figure 10: Sensitivity analysis of CohortNet on k and n in terms of AUC-PR on the MIMIC-III dataset.

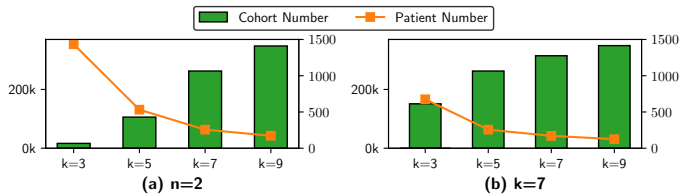


Figure 11: The cohort numbers and the average patient number in each cohort on the MIMIC-III dataset.

values of RR across all patients in a state-wise manner, as illustrated in Figure 9(a). Among these states, there is a specific state, namely S1, which is assigned to patients who have not undergone testing or recording for RR. Besides, we observe that different states of RR typically indicate different value ranges. For example, when a patient’s RR significantly exceeds the upper limit of the normal range (20 breaths per minute), it tends to fall into S6, indicating the patient’s relatively severe conditions.

Secondly, the transitions between the states of RR follow specific pathways, as shown in Figure 9(b). This diagram, based on all training samples, uses varying curve thicknesses to indicate the frequency of each state transition. It is evident that not all state pairs have direct transitions. For instance, there are one-way transitions from S2 to S3 and no direct transitions between S5 and S7. These validate our capability to effectively capture the evolution of RR.

Thirdly, the differentiation between a feature’s different states can be visualized by considering their co-existence with other features’ states. For instance, as shown in Figure 9(c), although S4 and S7 of RR exhibit similar average values, RR in S4 coexists with PH in S4 or S7, whereas RR in S7 never exhibit such coexistence. Delving into state correlations, CohortNet can capture more state distinctions with interpretable insights.

5.4.5 Feature-level Interaction Interpretation. In addition to learning from the above cohorts, CohortNet can capture diverse feature-level interactions through the Multi-channel Feature Learning Module. It determines the importance of these interactions, as measured by α in Eq. 2. Figure 8(e) illustrates an example where our model analyzes the feature RR, assigning higher significance to features with extremely abnormal values (such as BUN). Additionally, several other features (such as PCO2 and PIP) also receive elevated attention values, indicating their strong relevance to RR in the medical context. By integrating these crucial feature-level interactions, the representation learned in CohortNet for each feature can depict the feature’s various abnormal conditions, leading to improved performance and interpretability.

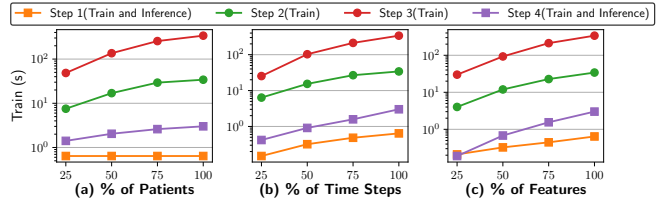


Figure 12: Scalability study of CohortNet on varying numbers of patients, time steps, features on the eICU dataset.

5.5 Sensitivity Study

In this experiment, we investigate the impact of two critical hyperparameters, k in Eq. 7 and n in Eq. 8, on the performance of CohortNet as illustrated in Figure 10. The k determines the number of states considered when analyzing each feature, while n affects the number of features used to define each cohort pattern.

Notably, CohortNet consistently exhibits improvements across different settings compared with the best-performing baselines, which affirms the effectiveness of learning cohorts. The highest AUC-PR value is achieved when $k = 7$ and $n = 2$. We also observe that moderate values of n and k are crucial for achieving substantial performance improvements. This is because both hyperparameters influence the trade-off between cohort granularity and the patient number associated with each cohort. As shown in Figure 11, smaller values of k or n result in more general cohorts with larger patient numbers. However, such cohorts possess coarser-grained states and fewer selected features, thus preserving less personalized information. Conversely, larger values of k or n result in cohorts that maintain more fine-grained details and tend to attain improved performance. Nonetheless, selecting overly high values of k or n may result in overfitting, and consequently, degrade performance.

5.6 Scalability Study

As detailed in Section 4.5, the time complexity of CohortNet is influenced by the number of patients, time steps, and features. To evaluate its scalability, we vary these factors in the largest eICU dataset and measure the runtime of each step, as shown in Figure 12.

In Step 1, CohortNet processes features across time steps to learn patients’ overall representations, with computation time increasing linearly with the number of features and time steps. In Steps 2 and 3, CohortNet utilizes the complete data across all patients and time steps to identify cohort patterns and learn cohort representations, respectively. As illustrated in Figure 12 (a) and (b), the runtime escalates more than linearly with an increase in the number of patients and time steps. This is because, with more patients and more time steps taken into consideration, CohortNet identifies more cohort patterns and requires a longer time to derive the representations of these cohorts. Additionally, an increase in the number of features leads to the discovery of more potential feature interactions during the cohort modeling process, subsequently extending the runtime for these two steps, as depicted in Figure 12 (c). Meanwhile, as CohortNet identifies an increased number of cohorts through these two steps, the computational demands also rise for training datasets or evaluating new patients in Step 4.

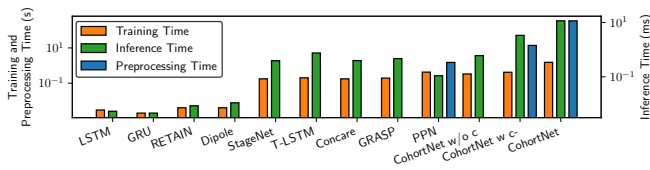


Figure 13: Runtime of models on the MIMIC-III dataset.

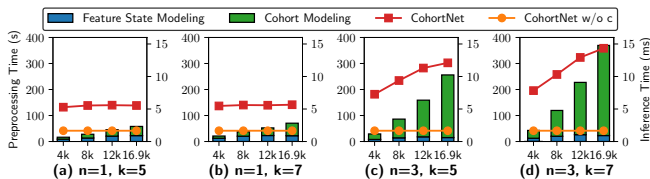


Figure 14: Efficiency analysis on k and n in terms of runtime on the MIMIC-III dataset with different sample sizes.

5.7 Efficiency Study

Efficiency comparison. We subsequently investigate the runtime of CohortNet in comparison with the baselines on the MIMIC-III dataset. The runtime metrics include the training time required for a batch of patients, the inference time for a new patient, and the preprocessing time required in Steps 2 and 3 for training. The comparative runtime results are depicted in Figure 13.

Among the baseline models, GRU and LSTM stand out for efficiently handling time series data, demonstrating superior performance in training and inference times. However, more complex models such as RETAIN and Dipole, which feature dual-layer or bidirectional GRU architectures, introduce additional computational overhead. Likewise, specialized designs like StageNet and T-LSTM, which are tailored to model disease stage progression and irregular time intervals respectively, necessitate increased runtime. Besides, compared with ConCare which also processes features separately, CohortNet w/o c exhibits relatively prolonged runtime due to its consideration of fine-grained feature interactions and trends. Regarding cohort modeling, GRASP learns exclusively from batches of patients, whereas PPN, CohortNet w c -, and CohortNet entail additional preprocessing time to learn typical patients or cohorts.

Effect of k and n on efficiency. We further evaluate the effect of k and n on our framework’s preprocessing efficiency for training and inference. During training, it encompasses all four steps, with k and n specifically influencing in the preprocesses of feature state learning and cohort representations learning in Steps 2 and 3. Figure 14 shows that as the number of patients increases, CohortNet demands more time since it needs to process all patients. Notably, when classifying each feature into fewer states (i.e., $k=5$) and exploring cohort patterns with less involved features (i.e., $n=1$), training time does not drastically increase with a larger sample size. It is attributed to the reduced cohort exploration spaces (i.e., lower k and n). Beyond a certain patient count, the growth of discovered cohorts is gradual, indicating that the identified cohort patterns are capable of encompassing future patients. However, CohortNet with a larger k needs a longer time to model features with more states. Similarly, a higher n indicates that more features will be considered

when CohortNet discovers cohort patterns. Consequently, as the training samples increase up to 16.9k, with a rise in both n and k , CohortNet discovers more cohorts, thereby elongating the time for subsequent patient retrieval and cohort learning.

During inference, altering sample size does not affect the time of CohortNet w/o c , which relies solely on individual patient data. In contrast, CohortNet takes longer for predicting since it leverages discovered cohorts when making decisions.

5.8 Discussions and Limitations

Selection of k and n . In our experimental setup, the k and n are fixed across all features determined by prediction performance. However, incorporating auxiliary information could improve their selections. For instance, in feature state modeling, the selection of k can be improved by considering feature characteristics such as missing rates and value ranges. For n in subsequent cohort pattern modeling, employing thresholds on α shows promise for automatically selecting n . Hence, adapting these parameters based on feature characteristics and interactions may improve final performance.

Time complexity. As analyzed in Section 4.5, utilizing more patients, features, and time steps indeed facilitates the discovery of more meaningful cohorts and improves prediction. However, this concurrently introduces a higher complexity and extends the time required for identifying, learning, and exploiting cohorts. To mitigate the computational burden, we could consider implementing advanced cohort filters and iterative cohort update strategies to shorten cohort learning time.

Cohort validation and knowledge exploitation. As shown in Section 5, discovered cohorts can significantly improve the prediction performance and provide medically meaningful insights, often aligning with previous medical studies. However, we acknowledge that not all high-performing cohorts demonstrate clear medical significance, necessitating further validation by clinicians.

6 CONCLUSIONS

In this paper, we present an interpretable healthcare analytics framework that facilitates effective cohort discovery with concrete feature patterns, which is a crucial task not achieved in prior studies. Its core component, CohortNet, comprises four modules designed for fine-grained patient representation learning considering individual feature trends and feature interactions, the auto-discovery of cohorts with medically interpretable patterns, comprehensive cohort representation learning through associated patient retrieval, and personalized cohort exploitation for enhanced healthcare analytics. Extensive experiments on three real-world EHR datasets show its superior performance, with AUC-PR improvements ranging from 2.8% to 4.1% compared to state-of-the-art baselines. These results confirm our framework’s effectiveness in cohort discovery, learning, and exploitation for enhancing prediction performance and providing interpretable insights. In future, we shall further collaborate with clinicians to validate our findings following medical practices.

ACKNOWLEDGMENTS

This research is supported by Singapore Ministry of Education Academic Research Fund Tier 3 under MOE’s official grant MOE2017-T3-1-007 and in part by NSF grants 2106176 and 2312931.

REFERENCES

- [1] ES Barker, RB Singer, JR Elkinton, JK Clark, et al. 1957. The renal response in man to acute experimental respiratory alkalosis and acidosis. *The Journal of clinical investigation* 36, 4 (1957), 515–529.
- [2] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. 2017. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 65–74.
- [3] Ricardo Cachucho, Siegfried Nijssen, and Arno Knobbe. 2017. Biclustering multivariate time series. In *Advances in Intelligent Data Analysis XVI: 16th International Symposium, IDA 2017, London, UK, October 26–28, 2017, Proceedings 16*. Springer, 27–39.
- [4] Qingpeng Cai, Kaiping Zheng, Beng Chin Ooi, Wei Wang, and Chang Yao. 2022. ELDA: Learning Explicit Dual-Interactions for Healthcare Analytics. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 393–406.
- [5] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in neural information processing systems* 29 (2016).
- [6] Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. 2018. Mime: Multi-level medical embedding of electronic health records for predictive healthcare. *Advances in neural information processing systems* 31 (2018).
- [7] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
- [8] Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning*. 233–240.
- [9] Inderjit S Dhillon. 2001. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 269–274.
- [10] Philip J Dorman, W James Sullivan, Robert F Pitts, et al. 1954. The renal response to acute respiratory acidosis. *The Journal of Clinical Investigation* 33, 1 (1954), 82–90.
- [11] Scott K Epstein and Nirupam Singh. 2001. Respiratory acidosis. *Respiratory care* 46, 4 (2001), 366–383.
- [12] Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M Glass, and Jimeng Sun. 2020. Stagenet: Stage-aware neural networks for health risk prediction. In *Proceedings of The Web Conference 2020*. 530–540.
- [13] Hrayr Harutyunyan, Hrant Khachatryan, David C. Kale, Greg Ver Steeg, and Aram Galstyan. 2019. Multitask learning and benchmarking with clinical time series data. *Scientific Data* 6, 1 (2019), 96. <https://doi.org/10.1038/s41597-019-0103-9>
- [14] Geoffrey Hinton. 2018. Deep learning—a technology with the potential to transform health care. *Jama* 320, 11 (2018), 1101–1102.
- [15] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [16] Yongsheng Huang, Xiaoyu Lyu, Dan Li, Lin Wang, Yujun Wang, Wenbin Zou, Yingxin Wei, and Xiaowei Wu. 2020. A cohort study of 676 patients indicates D-dimer is a critical risk factor for the mortality of COVID-19. *PLoS one* 15, 11 (2020), e0242045.
- [17] Dawei Jiang, Qingchao Cai, Gang Chen, HV Jagadish, Beng Chin Ooi, Kian-Lee Tan, and Anthony KH Tung. 2016. Cohort Query Processing. *Proceedings of the VLDB Endowment* 10, 1 (2016).
- [18] Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. 2023. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data* 10, 1 (2023), 1.
- [19] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3, 1 (2016), 1–9.
- [20] Stephen C Johnson. 1967. Hierarchical clustering schemes. *Psychometrika* 32, 3 (1967), 241–254.
- [21] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [22] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. 2017. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1903–1911.
- [23] Handong Ma, Jiahang Cao, Yuchen Fang, Weinan Zhang, Wenbo Sheng, Shaodian Zhang, and Yong Yu. 2022. Retrieval-Based Gradient Boosting Decision Trees for Disease Risk Assessment. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3468–3476.
- [24] Liantao Ma, Chaohe Zhang, Yasha Wang, Wenjie Ruan, Jiangtao Wang, Wen Tang, Xinyu Ma, Xin Gao, and Junyi Gao. 2020. Concare: Personalized clinical feature embedding via capturing the healthcare context. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 833–840.
- [25] James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. Oakland, CA, USA, 281–297.
- [26] Behrooz Omidvar-Tehrani, Sihem Amer-Yahia, and Laks VS Lakshmanan. 2018. Cohort representation and exploration. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 169–178.
- [27] Behrooz Omidvar-Tehrani, Sihem Amer-Yahia, and Laks VS Lakshmanan. 2020. Cohort analytics: efficiency and applicability. *The VLDB Journal* 29, 6 (2020), 1527–1550.
- [28] Beng Chin Ooi, Kian-Lee Tan, Sheng Wang, Wei Wang, Qingchao Cai, Gang Chen, Jinyang Gao, Zhaojing Luo, Anthony KH Tung, Yuan Wang, et al. 2015. SINGA: A distributed deep learning platform. In *Proceedings of the 23rd ACM international conference on Multimedia*. 685–688.
- [29] Cicero Pahins, Behrooz Omidvar-Tehrani, Sihem Amer-Yahia, Valérie Siroux, Jean-Louis Pepin, Jean-Christian Borel, and João Comba. 2019. COVIZ: a system for visual formation and exploration of patient cohorts. *Proceedings of the VLDB Endowment (PVLDB)* 12, 12 (2019), 1822–1825.
- [30] Maryam Panahiazar, Vahid Taslimitehrani, Naveen L Pereira, and Jyotishman Pathak. 2015. Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics. *Studies in health technology and informatics* 210 (2015), 369.
- [31] Eliana Pastor, Luca De Alfaro, and Elena Baralis. 2021. Looking for trouble: Analyzing classifier behavior via pattern divergence. In *Proceedings of the 2021 International Conference on Management of Data*. 1400–1412.
- [32] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019), 8026–8037.
- [33] Tom J Pollard, Alistair E W Johnson, Jesse D Raffa, Leo A Celi, Roger G Mark, and Omar Badawi. 2018. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific data* 5, 1 (2018), 1–13.
- [34] Jiarui Qin, Weinan Zhang, Rong Su, Zhirong Liu, Weiwen Liu, Ruiming Tang, Xiuqiang He, and Yong Yu. 2021. Retrieval & Interaction Machine for Tabular Data Prediction. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1379–1389.
- [35] Tabinda Sarwar, Sattar Seifollahi, Jeffrey Chan, Xiuzhen Zhang, Vural Aksakalli, Irene Lena Hudson, Karim Verspoor, and Lawrence Cavedon. 2023. The Secondary Use of Electronic Health Records for Data Mining: Data Characteristics and Challenges. *ACM Comput. Surv.* 55, 2 (2023), 33:1–33:40.
- [36] Seyedmostafa Sheikhalishahi, Vevake Balaraman, and Venet Osmani. 2019. Benchmarking machine learning models on multi-centre eICU critical care dataset. *arXiv:1910.00964* [cs.LG]
- [37] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Jing Gao, and Aidong Zhang. 2018. Deep patient similarity learning for personalized healthcare. *IEEE transactions on nanobioscience* 17, 3 (2018), 219–227.
- [38] Maxence Vandromme, Julie Jacques, Julien Taillard, Laetitia Jourdan, and Clarisse Dhaenens. 2020. A biclustering method for heterogeneous and temporal medical data. *IEEE Transactions on Knowledge and Data Engineering* 34, 2 (2020), 506–518.
- [39] Dawei Wang, Bo Hu, Chang Hu, Fangfang Zhu, Xing Liu, Jing Zhang, Binbin Wang, Hui Xiang, Zhenshun Cheng, Yong Xiong, et al. 2020. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *Jama* 323, 11 (2020), 1061–1069.
- [40] Wei Wang, Meihui Zhang, Gang Chen, HV Jagadish, Beng Chin Ooi, and Kian-Lee Tan. 2016. Database meets deep learning: Challenges and opportunities. *ACM Sigmod Record* 45, 2 (2016), 17–22.
- [41] Kazuhito Watanabe, Hsiang-Yun Wu, Yusuke Niibe, Shigeo Takahashi, and Issei Fujishiro. 2015. Biclustering multivariate data for correlated subspace mining. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 287–294.
- [42] LAWRENCE L WEED. 1968. Medical records that guide and teach (concluded). *Yearbook of Medical Informatics* 212 (1968), 1.
- [43] Jianfeng Wu, Jianqiang Huang, Guochao Zhu, Qiongya Wang, Qingquan Lv, Ying Huang, Yang Yu, Xiang Si, Hui Yi, Cuiqing Wang, et al. 2020. Elevation of blood glucose level predicts worse outcomes in hospitalized patients with COVID-19: a retrospective cohort study. *BMJ Open Diabetes Research and Care* 8, 1 (2020), e001476.
- [44] Zhongle Xie, Hongbin Ying, Cong Yue, Meihui Zhang, Gang Chen, and Beng Chin Ooi. 2020. Cool, a COhort OnLine analytical processing system. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 577–588.
- [45] Zhihao Yu, Chaohe Zhang, Yasha Wang, Wen Tang, Jiangtao Wang, and Liantao Ma. 2024. Predict and Interpret Health Risk using EHR through Typical Patients. *ICASSP (2024)*.
- [46] Chaohe Zhang, Xin Gao, Liantao Ma, Yasha Wang, Jiangtao Wang, and Wen Tang. 2021. GRASP: Generic Framework for Health Status Representation Learning Based on Incorporating Knowledge from Similar Patients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 715–723.
- [47] Yao Zhang, Yun Xiong, Yiheng Sun, Caihua Shan, Tian Lu, Hui Song, and Yangyong Zhu. 2022. RuDi: Explaining Behavior Sequence Models by Automatic

- Statistics Generation and Rule Distillation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2651–2660.
- [48] Kaiping Zheng, Shaofeng Cai, Horng Ruey Chua, Melanie Herschel, Meihui Zhang, and Beng Chin Ooi. 2022. DyHealth: Making Neural Networks Dynamic for Effective Healthcare Analytics. *Proc. VLDB Endow.* 15, 12 (2022), 3445–3458.
- [49] Kaiping Zheng, Shaofeng Cai, Horng Ruey Chua, Wei Wang, Kee Yuan Ngiam, and Beng Chin Ooi. 2020. TRACER: A Framework for Facilitating Accurate and Interpretable Analytics for High Stakes Applications. In *SIGMOD Conference*. ACM, 1747–1763.
- [50] Kaiping Zheng, Gang Chen, Melanie Herschel, Kee Yuan Ngiam, Beng Chin Ooi, and Jinyang Gao. 2021. PACE: Learning Effective Task Decomposition for Human-in-the-loop Healthcare Delivery. In *SIGMOD Conference*. ACM, 2156–2168.
- [51] Kaiping Zheng, Horng-Ruey Chua, Melanie Herschel, H. V. Jagadish, Beng Chin Ooi, and James Wei Luen Yip. 2024. Exploiting Negative Samples: A Catalyst for Cohort Discovery in Healthcare Analytics. In *International Conference on Machine Learning (Proceedings of Machine Learning Research)*. PMLR.
- [52] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The lancet* 395, 10229 (2020), 1054–1062.
- [53] Zihao Zhu, Changchang Yin, Buyue Qian, Yu Cheng, Jishang Wei, and Fei Wang. 2016. Measuring patient similarities via a deep architecture with medical concept embedding. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 749–758.